JOURNAL OF BACTERIOLOGY, Apr. 2009, p. 2474–2484 Vol. 191, No. 8
0021-9193/09/$08.00+0 doi:10.1128/JB.01786-08

Phylogeography of *Francisella tularensis*: Global Expansion of a Highly Fit Clone▽†

Amy J. Vogler,[1] Dawn Birdsell,[1] Lance B. Price,[2] Jolene R. Bowers,[2] Stephen M. Beckstrom-Sternberg,[1,2] Raymond K. Auerbach,[1]‡ James S. Beckstrom-Sternberg,[1] Anders Johansson,[3] Ashley Clare,[1] Jordan L. Buchhagen,[1] Jeannine M. Petersen,[4] Talima Pearson,[1] Josée Vaissaire,[5] Michael P. Dempsey,[6] Paul Foxall,[7] David M. Engelthaler,[2] David M. Wagner,[1] and Paul Keim[1,2]*

Center for Microbial Genetics and Genomics, Northern Arizona University, Flagstaff, Arizona 86011-4073[1]; Translational Genomics Research Institute, Phoenix, Arizona 85004[2]; Department of Clinical Microbiology, Infectious Diseases and Bacteriology, Umeå University, SE 901 85 Umeå, and Division of CBRN Defence and Security, Swedish Defence Research Agency, SE 901 82 Umeå, Sweden[3]; Centers for Disease Control and Prevention, Fort Collins, Colorado 80521[4]; Agence Française de Sécurité Sanitaire des Aliments, Laboratoire d'Etudes et de Recherches en Pathologie Animale et Zoonoses, 94700 Maison-Alfort, France[5]; Division of Microbiology, Armed Forces Institute of Pathology, Washington, D.C. 20306[6]; and Affymetrix, Inc., Santa Clara, California 95051[7]

Received 19 December 2008/Accepted 10 February 2009

***Francisella tularensis* contains several highly pathogenic subspecies, including *Francisella tularensis* subsp. *holarctica*, whose distribution is circumpolar in the northern hemisphere. The phylogeography of these subspecies and their subclades was examined using whole-genome single nucleotide polymorphism (SNP) analysis, high-density microarray SNP genotyping, and real-time-PCR-based canonical SNP (canSNP) assays. Almost 30,000 SNPs were identified among 13 whole genomes for phylogenetic analysis. We selected 1,655 SNPs to genotype 95 isolates on a high-density microarray platform. Finally, 23 clade- and subclade-specific canSNPs were identified and used to genotype 496 isolates to establish global geographic genetic patterns. We confirm previous findings concerning the four subspecies and two *Francisella tularensis* subsp. *tularensis* subpopulations and identify additional structure within these groups. We identify 11 subclades within *F. tularensis* subsp. *holarctica*, including a new, genetically distinct subclade that appears intermediate between Japanese *F. tularensis* subsp. *holarctica* isolates and the common *F. tularensis* subsp. *holarctica* isolates associated with the radiation event (the B radiation) wherein this subspecies spread throughout the northern hemisphere. Phylogenetic analyses suggest a North American origin for this B-radiation clade and multiple dispersal events between North America and Eurasia. These findings indicate a complex transmission history for *F. tularensis* subsp. *holarctica*.**

Francisella tularensis, the etiologic agent of tularemia, is a facultative intracellular pathogen known to infect numerous animal species, including humans (17), and is a category A select agent (4). *F. tularensis* is currently divided into three subspecies, *Francisella tularensis* subsp. *tularensis*, *Francisella tularensis* subsp. *holarctica*, and *Francisella tularensis* subsp. *mediasiatica*, which differ in their pathogenicities and geographic distributions (19). *Francisella novicida*, officially recognized as another species within the *Francisella* genus, is informally considered the fourth subspecies of *F. tularensis* on the basis of its similarity in DNA hybridization experiments (19) and will be treated as such in this publication. *F. tularensis* subsp. *tularensis* causes life-threatening type A tularemia and is found only in North America. *F. tularensis* subsp. *holarctica* causes less-severe type B tularemia and is found throughout the northern hemisphere. *F. tularensis* subsp. *mediasiatica* has virulence similar to that of *F. tularensis* subsp. *holarctica* but is geographically restricted, having been isolated only from central Asia (19). *Francisella tularensis* subsp. *novicida* is the least virulent subspecies, having been isolated only rarely in North America (19) and once in Australia (35).

The four subspecies differ in their genetic diversity. The rare *F. tularensis* subsp. *mediasiatica* and *F. tularensis* subsp. *novicida* subspecies appear to contain significant diversity (18), but due to the scarcity of samples, their true genetic diversity remains unknown. The two dominant and highly pathogenic subspecies, *F. tularensis* subsp. *tularensis* and *F. tularensis* subsp. *holarctica*, differ substantially in their levels of genetic diversity. *F. tularensis* subsp. *tularensis* has two distinct, genetically differentiated subpopulations (A.I and A.II), with separate geographic distributions (11, 18, 19, 29, 30). This division in *F. tularensis* subsp. *tularensis* is apparent via a variety of molecular typing methods, including whole-genome single nucleotide polymorphism (SNP) analysis (19), multilocus sequence typing (30), ribotyping (12), regional difference (RD)

* Corresponding author. Mailing address: Center for Microbial Genetics and Genomics, Northern Arizona University, P.O. Box 4073, Flagstaff, AZ 86011. Phone: (928) 523-1078. Fax: (928) 523-4015. E-mail: Paul.Keim@nau.edu.

† Supplemental material for this article may be found at http://jb.asm.org/.

‡ Present address: Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT.

▽ Published ahead of print on 27 February 2009.

analysis (8), pulsed-field gel electrophoresis (12, 29), amplified fragment length polymorphism (12, 13), canonical insertion-deletions (22), and multilocus variable-number tandem repeat (VNTR) analysis (MLVA) (11, 18). These subpopulations are correlated with different host and vector distributions (11, 19) and may differ in virulence (29). In addition, subpopulations A.I and A.II have been shown to contain substantial genetic diversity via the relatively high-resolution molecular methods of pulsed-field gel electrophoresis (29) and MLVA (11, 18). In contrast, very little genetic diversity has been identified within *F. tularensis* subsp. *holarctica* with any of the molecular methods outlined above. This lack of diversity, combined with *F. tularensis* subsp. *holarctica*'s wide geographic distribution, has led several authors to suggest that this subspecies recently experienced a genetic bottleneck and expanded across the northern hemisphere (8, 11, 18, 19).

As a consequence, the lack of genetic diversity within *F. tularensis* subsp. *holarctica* has made defining population structure within this subspecies especially problematic. Although all evidence points to a recent global expansion of this subspecies, there is little information on this expansion. MLVA provides the greatest discrimination among *F. tularensis* subsp. *holarctica* isolates (18, 19), but highly mutable VNTR markers can be compromised for phylogenetic analyses due to the likelihood of convergent evolution and the resulting homoplasy (20). Indeed, MLVA of global *F. tularensis* subsp. *holarctica* isolates has identified a few cases where North American and Scandinavian isolates are similar, although the majority of the data indicated that these two populations are distinct (18). This could be due to the hypervariability and fortuitous convergent evolution of the markers used (homoplasy) or may represent multiple dispersals of *F. tularensis* to North America, Scandinavia, or both. A set of evolutionarily stable, nonhomoplastic molecular markers would provide the means for addressing such questions by defining deeper population structure within *F. tularensis* and, especially, within *F. tularensis* subsp. *holarctica*.

Whole-genome SNP analysis has been shown to be remarkably effective at defining population structure among highly clonal bacterial pathogens lacking much genetic diversity (1, 20, 26, 31), as long as the discovery bias inherent to this method is taken into consideration (26). Because of the vast amount of sequence coverage, this method has the capacity to discover thousands of SNPs, even among very recently emerged genetically homogeneous organisms (20, 26). Once an accurate population structure has been defined using these SNPs, canonical SNPs (canSNPs) that define each branch in the phylogeny (20), whether species specific (9, 10), major lineage specific (31), or strain specific (32, 34), can be selected. This approach maximizes the information obtained, while limiting the number of assays, by eliminating redundancy and can be combined, in a stepwise hierarchical fashion, with more-variable molecular markers to provide highly accurate and highly discriminating phylogenetic analyses (20). The ever-expanding availability of whole-genome sequences for comparison has made this approach feasible for a number of important bacterial pathogens, including *F. tularensis*.

In this study, we further investigated the phylogenetic structure of *F. tularensis* via three distinct SNP analyses that together maximized our genomic and isolate coverage: whole-genome SNP analysis, molecular inversion probe (MIP) SNP analysis, and canSNP analysis (Table 1). We discovered SNPs among 13 *F. tularensis* genomes and used them to construct a highly accurate phylogeny for *F. tularensis*. We used high-throughput MIP microarray technology to genotype 1,655 SNPs across 95 genetically and geographically diverse *F. tularensis* isolates to identify additional phylogenetic structure and designate new canonical groups within *F. tularensis*, with an emphasis on *F. tularensis* subsp. *holarctica*. Finally, we defined canSNPs for these groups and screened them across 496 genetically and geographically diverse *F. tularensis* isolates to identify the worldwide distribution of these groups. Altogether, our phylogenetic and geographic distribution analyses represent the most comprehensive description to date of the worldwide diversity and historical spread of *F. tularensis*.

TABLE 1. *F. tularensis* SNP analysis summary[a]

Analysis	No. of SNP loci	No. of isolates
Whole-genome SNP	29,774	13
MIP SNP	1,655	95
canSNP	23	496

[a] This table lists the number of SNP loci and the number of isolates analyzed in each of three analyses. For details on the selection of the SNP loci used in each analysis, see Materials and Methods.

MATERIALS AND METHODS

A summary flow chart of the three principal SNP analyses (whole-genome SNP analysis, MIP SNP analysis, and canSNP analysis) is presented in Fig. 1.

Whole-genome SNP discovery and analysis. An in-house pipeline (2) was used to discover SNPs among 13 available *F. tularensis* whole-genome sequences: those for *F. tularensis* subsp. *holarctica* LVS (NC_007880), *F. tularensis* subsp. *holarctica* FTNF002-00 (NC_009749), *F. tularensis* subsp. *holarctica* OSU18 (NC_008369) (27), *F. tularensis* subsp. *holarctica* OR96-0246 (Baylor College of Medicine Human Genome Sequencing Center), *F. tularensis* subsp. *holarctica* 257 (NZ_AAUD00000000), *F. tularensis* subsp. *holarctica* FSC 200 (NZ_AASP00000000), *F. tularensis* subsp. *holarctica* FSC 022 (NZ_AAYD00000000), *F. tularensis* subsp. *tularensis* SCHU S4 (NC_006570) (21), *F. tularensis* subsp. *tularensis* FSC 033 (NZ_AAYE00000000), *F. tularensis* subsp. *tularensis* ATCC 6223 (Baylor College of Medicine Human Genome Sequencing Center), *F. tularensis* subsp. *tularensis* WY96-3418 (NC_009257) (3), *F. tularensis* subsp. *mediasiatica* FSC 147 (NC_010677), and *F. tularensis* subsp. *novicida* U112 (NC_008601) (28). In brief, the in-house pipeline utilized MUMmer to perform pairwise local alignments of 200-bp segments on a sliding window. Since this analysis compared sequence segments and did not require synteny, recombination was not a factor in SNP discovery. The pipeline then utilized a series of custom Perl and Java scripts for tabulating the identified SNPs in accordance with certain criteria. Specifically, a SNP was required to be from a region shared across all 13 *F. tularensis* genomes and to have at least 30 bp of unvarying flanking sequence on each side of the SNP. Thus, this analysis included SNPs from all shared regions of the chromosome, both intergenic and intragenic (see Fig. S1 in the supplemental material). Apparent tri-state SNPs were removed from the analysis on the basis of the assumption that they were most likely due to sequencing errors (Fig. 1), though this assumption was not confirmed. Homoplastic SNPs were included in the analysis, though the very low number of these SNPs (250 out of the 29,774 SNPs used in the phylogenetic analysis) (Fig. 1) supported the clonal nature of *F. tularensis* and indicated a lack of recombination in this species. The software package PAUP 4.0b10 (D. Swofford, Sinauer Associates, Inc., Sunderland, MA) was used to construct a whole-genome SNP phylogeny (Fig. 2).

MIP SNP selection and analysis. At the time of MIP SNP selection, only eight *F. tularensis* genome sequences were available for SNP identification: those for LVS, FTNF002-00, OSU18, OR96-0246, SCHU S4, ATCC 6223, WY96-3418, and FSC 147. The same in-house pipeline (2) was used to discover SNPs among these eight genomes (Fig. 1). SNPs representing phylogenetic branch points of

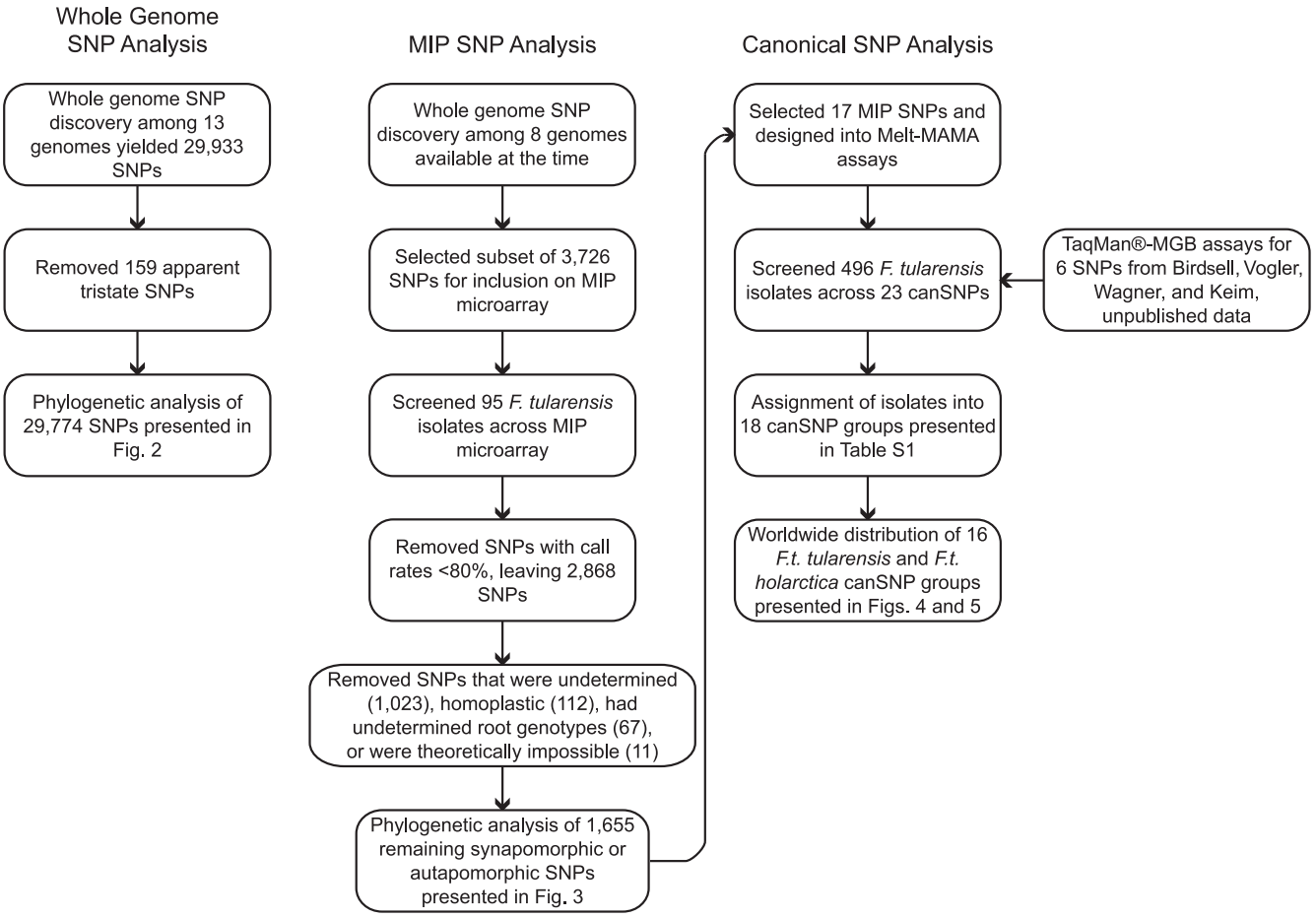

FIG. 1. Summary flow chart of three SNP analyses used. The methods for the three principal SNP analyses (whole-genome SNP analysis, MIP SNP analysis, and canSNP analysis) are presented.

interest were selected. These branches included A versus B, A.I versus A.II, A.I versus B, and A.II versus B as well as other comparisons within each group. Several potential strain-specific SNPs were also included. Synonymous, nonsynonymous, and intergenic SNPs (relative to the SCHU S4 sequence annotation) were included on the MIP microarray. Each SNP was extracted with corresponding flanking sequence by using a Perl script, and these sequences were submitted to Affymetrix (Santa Clara, CA) for quality control and probe design.

A total of 3,726 putative SNPs were selected for inclusion in the MIP probe panel on the basis of their potential for identifying new subclades and the available space on the microarray (Fig. 1). MIP assay analysis was performed as

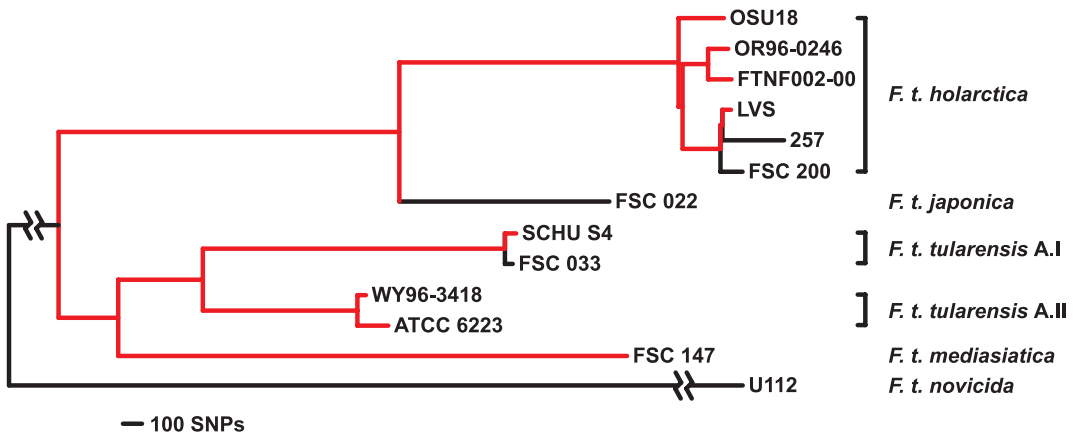

FIG. 2. Whole-genome SNP phylogeny of 13 *F. tularensis* strains. Maximum parsimony was used to construct this phylogeny from 29,774 SNPs discovered among the whole-genome sequences of these 13 strains. Red branches indicate the branches along which SNPs were discovered for the MIP SNP analysis, as only eight whole-genome sequences were available for SNP discovery when the MIP SNPs were selected. This analysis was highly robust due to the large number of characters and the very high consistency index of >0.99.

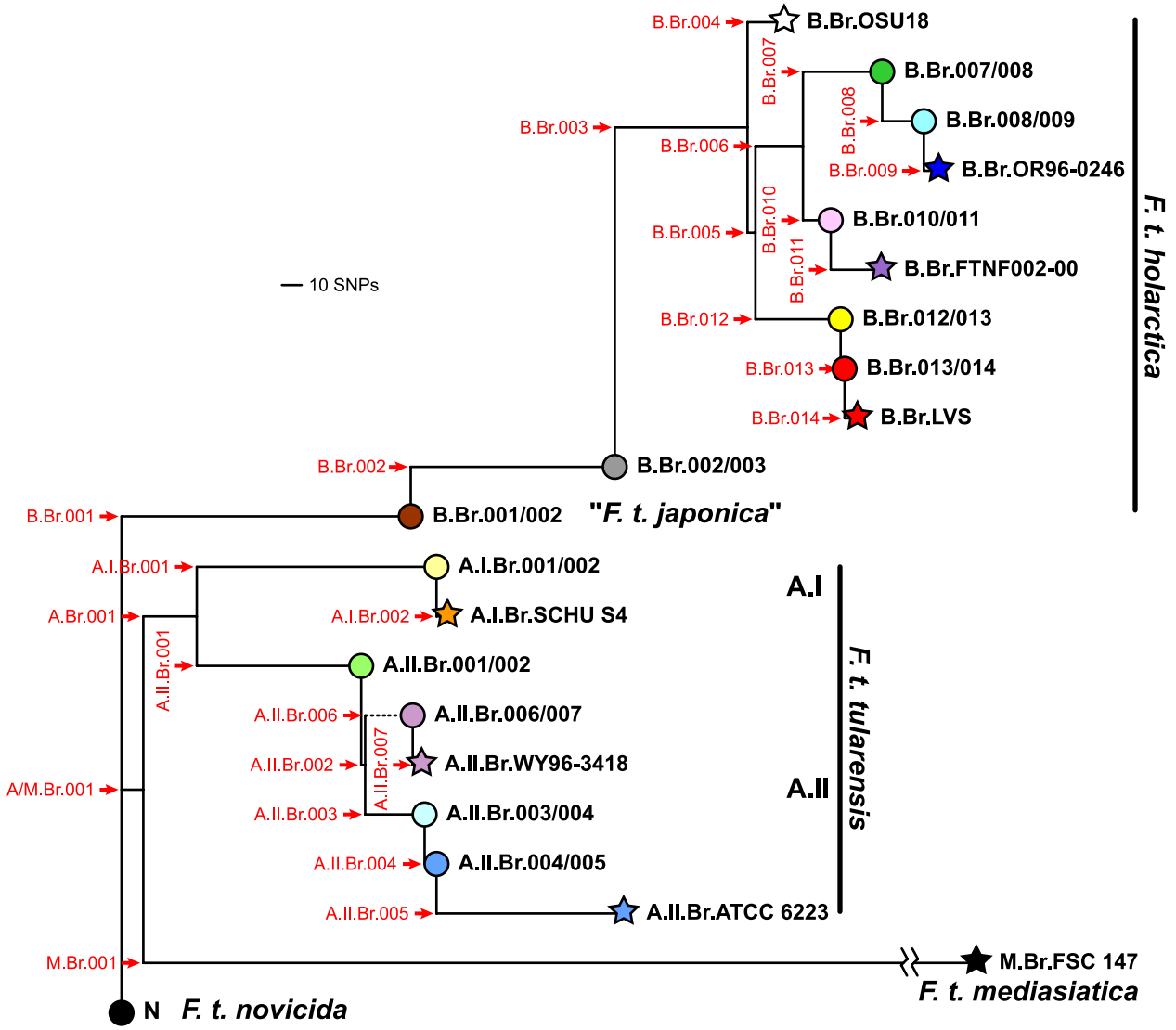

FIG. 3. MIP SNP phylogeny indicating canSNPs, clades, and subclades. Stars indicate terminal subclades defined by one of the eight *F. tularensis* genomes used for MIP SNP discovery. Circles represent collapsed branch points along the lineages that contain subgroups of isolates. These subclades are named for the two flanking canSNPs. canSNP names and positions are indicated in red. Branch A.II.Br.006 has a branch length of 1 SNP but is presented as a longer, dashed line to avoid obscuring other parts of the tree.

previously described (15, 16) on whole-genome amplification (WGA) products (Qiagen, Valencia, CA) from 95 genetically and geographically diverse *F. tularensis* isolates (Fig. 1; also see Table S1 in the supplemental material). Briefly, 4.5 μl of a 1/10 dilution of WGA product was added to the MIP reaction mixture containing the 3,726 *F. tularensis* probes. Probes were annealed to target DNA, starting at 95°C and ramping slowly to 58°C (−0.2°C every 5 min), and then held at 58°C for a total annealing time of 16 to 24 h. The mixture was then split into four tubes for nucleotide-specific gap filling, ligation, exonuclease digestion of nonligated probes, probe cleavage, amplification, and labeling according to the manufacturer's specifications. Labeled probes were repooled and hybridized to the tag arrays overnight at 39°C. Tag arrays were washed and stained according to the manufacturer's specifications and scanned using a GeneChip AT charge-coupled-device imager, and the data were analyzed using Affymetrix GeneChip targeted genotyping analysis software. As this system was designed for diploid organisms, the resulting genotype data consisted of "homozygous" SNP calls for either of two possible states, undetermined calls and "heterozygous" SNP calls. "Heterozygous" results were treated as undetermined for this haploid organism.

Genotype data from polymorphic MIP SNPs with ≥80% call data across the 95-isolate screening panel were used to construct a phylogeny detailing subclades discovered among the screened isolates (Fig. 1 and 3). Isolates with missing data were assigned SNP states for those SNPs with undetermined genotype calls. These SNP calls were consistent with the SNP states observed in other redundant SNP markers. This approach was justified given the very low homoplasy observed in the whole-genome sequence SNP analysis and the clonality of *F. tularensis*. This approach ensures that the overall length of the resulting phylogeny is conserved and not inflated due to missing data being assigned values that would increase the homoplasy. The phylogeny was constructed using the software package PAUP 4.0b10 (D. Swofford, Sinauer Associates, Inc., Sunderland, MA). MIP SNPs within this subset that exhibited homoplasy or theoretically impossible derived states (i.e., unique derived states in nondiscovery strains) (26, 36) were eliminated from the phylogenetic analysis given their unlikelihood and the impracticality of confirming the results via another method. Undetermined and synapomorphic MIP SNPs for which ancestral and derived groups could not be identified, due to undetermined results in root isolates, were also eliminated from the phylogenetic analysis (Fig. 1). Undetermined MIP SNPs were defined as SNPs with no observed derived states across the 95 screened isolates (derived states could have existed among isolates with missing data) or SNPs for which a large percentage of likely "derived isolates" were missing data. These steps were taken to ensure the most conservative approach for subclade discovery. Branch and subclade naming followed the approach suggested by Van Ert et al. (31), in

which terminal groups containing discovery strains are referred to as subclades named for the discovery genome and intervening nodes are referred to as subclades named for their neighboring branches. Branches are named based upon their major group affiliation (A, A.I, A.II, B, M, or N for the different subspecies and two major subpopulations of *F. tularensis* subsp. *tularensis*) and an arbitrary three-digit number.

canSNP selection and analysis. Seventeen MIP SNPs defining the subclades discovered using the MIP assay were selected as canSNPs and developed into high-throughput SNP genotyping assays that we term Melt-MAMAs (Fig. 1). These assays utilize allele-specific mismatch amplification mutation assay (MAMA) primers (5) coupled with GC- and T-rich primer tails, SYBR green dye, and melting curve analysis for SNP genotyping (25). These assays and an additional six TaqMan minor groove binding canSNP allelic discrimination assays for major groups within *F. tularensis* (D. Birdsell, A. J. Vogler, D. M. Wagner, and P. Keim, unpublished data) were screened across 496 genetically and geographically diverse *F. tularensis* isolates (Fig. 1; also see Table S1 in the supplemental material), including the 95-strain MIP screening panel. Melt-MAMA primers were designed using Primer Express 3.0 software (Applied Biosystems, Foster City, CA) and the guidelines presented by Li et al. (23). Derived allele primers contained a GC-rich tail, while ancestral allele primers contained a T-rich tail, giving the two allele-specific PCR products easily distinguishable melting temperatures that differed by >3°C. The genome locations for the 17 chosen canSNPs and the primer sequences for their Melt-MAMA primers can be found in Table S2 in the supplemental material. Each 5-μl Melt-MAMA reaction mixture contained 1× SYBR green PCR master mix, derived and ancestral allele-specific MAMA primers, a common reverse primer (for primer concentrations, see Table S2 in the supplemental material), and 1 μl of a diluted DNA template. DNA templates were diluted 1/10 for heat soak preparations or 1/50 for WGA products (Qiagen, Valencia, CA). All Melt-MAMAs were performed using the same thermal cycling parameters on an Applied Biosystems 7900HT Fast real-time PCR system with SDS v2.3 software. Reactions were first raised to 50°C for 2 min to activate the uracil glycolase, then raised to 95°C for 10 min to denature the DNA, and then cycled at 95°C for 15 s and 55°C for 1 min for 33 cycles. Melting curve analysis was performed immediately following the completion of the PCR by ramping from 60 to 95°C in 0.2°C/min increments and recording the fluorescence. The negative first derivative of these fluorescent readings was plotted, yielding a graph with a peak at the melting temperature of the PCR product, which was used for genotyping purposes.

MLVA. All 496 *F. tularensis* isolates were also genotyped using an 11-marker MLVA system (33). This was done in order to determine the level of genetic diversity within each identified subclade.

Geographic distribution of clades and subclades. We mapped the worldwide distributions of the 16 synapomorphic *F. tularensis* subsp. *tularensis* and *F. tularensis* subsp. *holarctica* subclades that we identified to determine their genetic geographic patterns (Fig. 1). This was done using color codes for each subclade and corresponding pie charts for each major geographic region that indicated both the number of isolates from a given region and the proportion of each subclade found within that region.

RESULTS

We used three different SNP analyses to understand the global population structure of *F. tularensis* subsp. *holarctica*: whole-genome SNP analysis, MIP SNP analysis, and canSNP analysis (Table 1). We present a highly accurate phylogeny for *F. tularensis* based upon 29,774 SNPs discovered among 13 *F. tularensis* whole-genome sequences. We used 1,655 SNPs to genotype 95 *F. tularensis* isolates by using MIP technology and identified 18 new synapomorphic and 3 new autapomorphic (genome-specific) canSNP groups (subclades) for *F. tularensis*, 11 of which are within *F. tularensis* subsp. *holarctica*. We defined canSNPs for these 18 synapomorphic subclades and used them to determine the subclades for 496 *F. tularensis* isolates. We mapped the geographic distributions of 6 *F. tularensis* subsp. *tularensis* subclades across North America and 10 *F. tularensis* subsp. *holarctica* subclades across the world. We used these geographic distributions to suggest potential historical transmission routes for *F. tularensis*.

Whole-genome SNP analysis. Comparisons among 13 *F. tularensis* whole-genome sequences led to the discovery of 29,933 SNPs within this species (Fig. 1). We used 29,774 of these SNPs to construct a detailed and robust phylogenetic hypothesis for these diverse strains (Fig. 1 and 2). Consistent with a previous analysis (19), *F. tularensis* subsp. *tularensis* and *F. tularensis* subsp. *mediasiatica* are sister taxa that split at a deeper node into their own lineage whereas *F. tularensis* subsp. *holarctica* occupies a lineage by itself. The A.I and A.II subpopulation division within *F. tularensis* subsp. *tularensis* is clearly visible, with limited diversity between the two strains included in each of these subpopulations. The single Japanese *F. tularensis* subsp. *holarctica* genome (FSC 022) is genetically distinct from the other *F. tularensis* subsp. *holarctica* strains and diverged from the lineage leading to the remaining *F. tularensis* subsp. *holarctica* isolates (Fig. 2). This is consistent with previous findings of its divergence from other *F. tularensis* subsp. *holarctica* isolates (8, 18, 30); this supports the suggestion that Japanese *F. tularensis* subsp. *holarctica* isolates may constitute a separate subspecies, identified as *Francisella tularensis* subsp. *japonica* (24). *F. tularensis* subsp. *novicida* is genetically distant from the other *F. tularensis* subspecies and was used only to root the tree in accordance with previous analyses (19).

MIP SNP analysis. In order to further refine the phylogenetic topology of closely related strains, we used 1,655 SNPs to genotype 95 genetically and geographically diverse *F. tularensis* isolates by using a MIP assay. These SNPs represented a subset of the total SNPs discovered among eight *F. tularensis* whole-genome sequences that were available at the time of MIP SNP discovery. This is one of the first reports of MIP technology applied to pathogens, indicating that this technology is a viable approach for dense SNP genotyping of prokaryotic populations and for the use of WGA templates.

Discovery bias is a major consideration when limited but comprehensive whole-genome sequence discovery is coupled with high-capacity genotyping methods. Whole-genome sequence SNP discovery will discover only SNPs that occur on the evolutionary path between any two compared genomes (26, 36). Hence, our MIP SNP tree, based upon SNPs discovered via pairwise comparisons among eight discovery genomes, represents a combination of 28 linear (pairwise) phylogenies and possesses eight termini (Fig. 3). We also identify 13 intervening nodes (Fig. 3), which group isolates together at the point where they started their own evolutionary pathways. These secondary branches are collapsed because their independent evolutionary history was not captured during the MIP SNP discovery process (Fig. 2 and 3) (26, 36). As expected, the MIP SNP tree topology parallels that of the whole-genome sequence SNP tree seen in Fig. 2, except for branch lengths (Fig. 2 and 3). Branch lengths in Fig. 2 are highly accurate since they are based upon all SNPs discovered among the 13 strains. Because only a subset of the whole-genome SNPs were screened using the MIP assay (Fig. 1), the MIP SNP tree contains inaccurate branch lengths. However, the relative positioning of subclades and overall topology are highly accurate.

Novel genomes can easily be placed onto the phylogenetic tree by in silico genotyping of the SNPs used in this analysis. For example, FSC 022, the Japanese *F. tularensis* subsp. *holarctica* whole-genome sequence that was publicly released after the MIP design and not included in the MIP SNP discovery

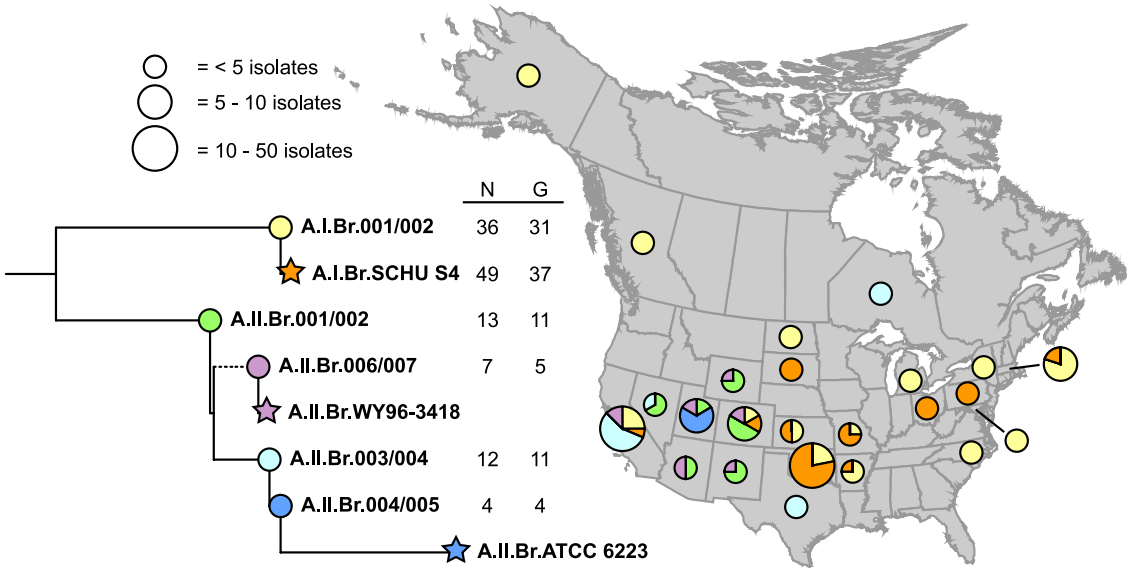

FIG. 4. Distribution of *F. tularensis* subsp. *tularensis* subclades throughout North America. The *F. tularensis* subsp. *tularensis* portion of the canSNP phylogeny from Fig. 3 is presented along with a map indicating the frequencies and geographic distribution of *F. tularensis* subsp. *tularensis* subclades across North America. The number of isolates (N) and number of MLVA genotypes (G) within each subclade are indicated. Subclades are color coded to facilitate mapping. For the purposes of mapping and determining isolate and MLVA genotype totals, strains belonging in strain-specific subclades (A.II.Br.ATCC 6223 and A.II.Br.WY96-3418) are included within the subclade immediately basal to their actual subclade. Colors within the mapped pie charts correspond to the subclade color designations. For explanation of stars and circles, see the legend to Fig. 3.

process, is a member of subclade B.Br.001/002 on the MIP SNP tree (Fig. 3; also see Table S1 in the supplemental material). This is the subclade representing the collapsed Japanese *F. tularensis* subsp. *holarctica* (*F. tularensis* subsp. *japonica*) branch that can be observed in Fig. 2. Similarly, the *F. tularensis* subsp. *novicida* U112 genome is categorized with the *F. tularensis* subsp. *novicida* subclade (N) on the MIP SNP tree (Fig. 3; also see Table S1 in the supplemental material). Three other examples are the FSC 200 and 257 strain genomes, which can be categorized as members of subclade B.Br.013/014, and the FSC 033 strain genome, which belongs to subclade A.I.Br.001/002 (Fig. 3). When additional *F. tularensis* genome sequences become available, they could be placed into this phylogenetic structure in a similar manner.

Three of the eight terminal subclades, B.Br.LVS, A.II.Br. WY96-3418, and A.II.Br.ATCC 6223, were found to contain only the corresponding discovery strain after the MIP SNP analysis across the 95 genetically and geographically diverse strains (Fig. 3; also see Table S1 in the supplemental material), indicating that the SNPs identifying these terminal groups were likely strain specific. As such, SNPs for these branches (A.II. Br.005, A.II.Br.007, and B.Br.014) (Fig. 3) were not converted into high-throughput assays or screened against the larger, 496-strain panel. Potential strain-specific SNPs were not identified for terminal subclades B.Br.OR96-0246, B.Br.FTNF002-00, and M.Br.FSC 147, since the corresponding discovery strain was not included in the MIP assay screening panel (Fig. 3; also see Table S1 in the supplemental material). Although the corresponding discovery strain was included in the MIP assay screening panel for terminal subclades B.Br.OSU18 and A.I.Br.SCHU S4 (Fig. 3; also see Table S1 in the supplemental material), the use of an incomplete sequence and the lack of a second A.I strain at the time of MIP SNP discovery, respectively, may have hampered the discovery of potential strain-specific SNPs for these two strains.

canSNP and MLVA genotyping analyses. Figures 4 and 5 graphically depict the distribution of 484 of the 496 isolates in 16 subclades of *F. tularensis* subsp. *tularensis* and *F. tularensis* subsp. *holarctica* identified in the MIP SNP analysis (Fig. 4 and 5, columns N). The remaining 12 isolates were either *F. tularensis* subsp. *mediasiatica* or *F. tularensis* subsp. *novicida*, were members of subclade M.Br.FSC 147 or N, respectively (see Table S1 in the supplemental material), and are not depicted in Fig. 4 or 5. Subclade placement for the 496 isolates was determined by their canSNP genotypes across 23 canSNPs (see Table S1 in the supplemental material). canSNP genotypes for the 95 isolates screened using the MIP assay were consistent with the available MIP assay results for those SNPs, although some missing MIP assay data prevented a full comparison. The position of each of the 23 canSNPs is illustrated in Fig. 3, and the canSNP genotype for each of the 18 synapomorphic subclades is depicted in Table 2. canSNPs for branches A/M.Br.001, M.Br.001, A.Br.001, A.I.Br.001, A.II.Br.001, and B.Br.001 (Fig. 3) were identified elsewhere and were genotyped using TaqMan minor groove binding allelic discrimination assays (D. Birdsell, A. J. Vogler, D. M. Wagner, and P. Keim, unpublished data). canSNPs for branches A.I.Br.002, A.II.Br.002-004, A.II.Br.006, and B.Br.002-013 (Fig. 3) were selected from among the MIP SNPs and were genotyped using the Melt-MAMAs described here (see Materials and Methods). As mentioned previously, canSNPs for branches A.II.Br.005, A.II.Br.007, and B.Br.014 (Fig. 3) were not selected or screened, due to the probable strain-specific nature of these SNPs.

MLVA genotyping analysis revealed various levels of genetic diversity among the subclades (Fig. 4 and 5, columns G). Consistent with previous MLVAs (18), *F. tularensis* subsp. *tularen-*

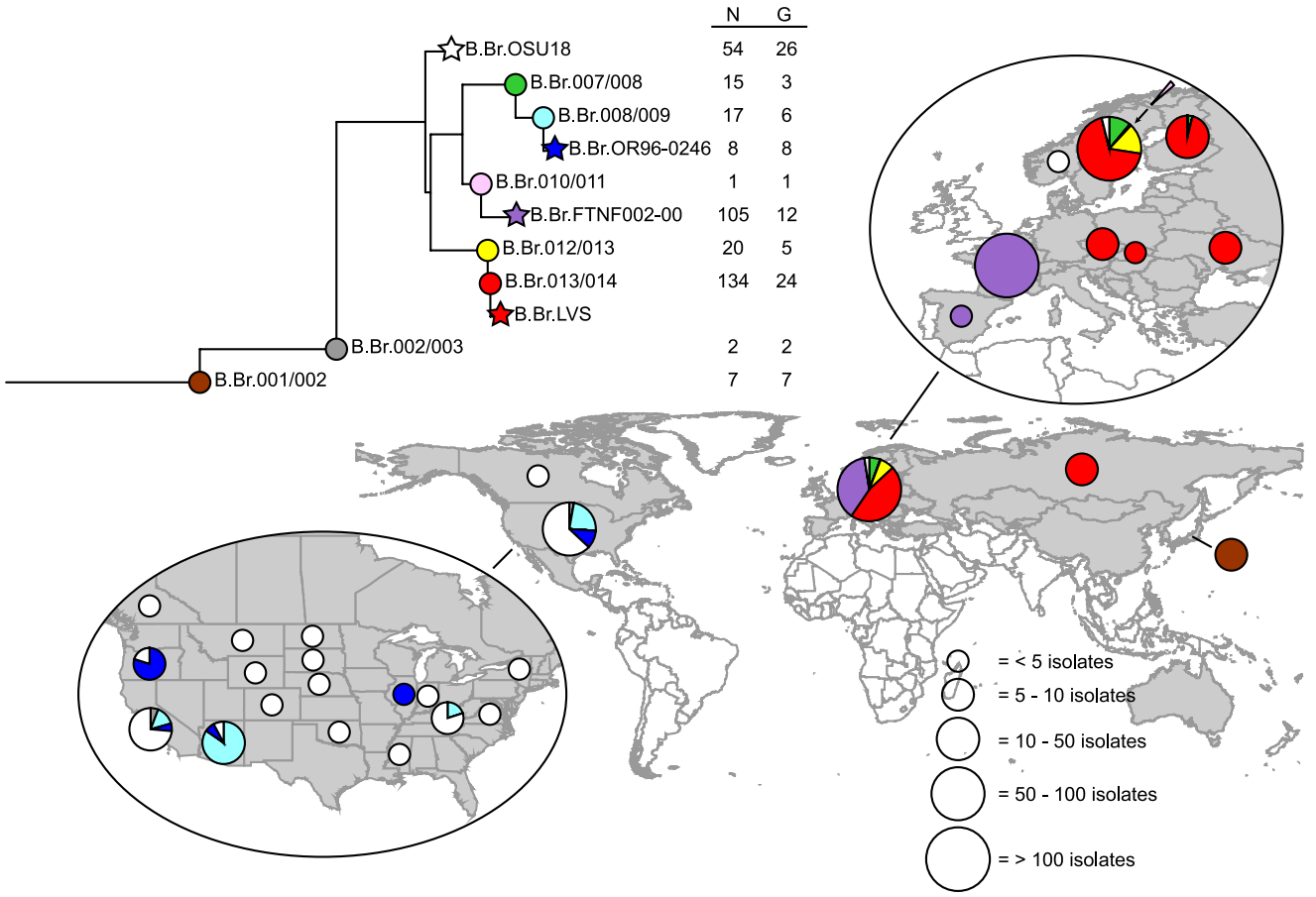

FIG. 5. Worldwide distribution of *F. tularensis* subsp. *holarctica* subclades. The *F. tularensis* subsp. *holarctica* portion of the canSNP phylogeny from Fig. 3 is presented along with a map indicating the frequencies and geographic distribution of *F. tularensis* subsp. *holarctica* subclades throughout the world. The number of isolates (N) and number of distinct MLVA genotypes (G) within each subclade are indicated. Subclades are color coded to facilitate mapping. For the purposes of mapping and determining isolate and MLVA genotype totals, the strain belonging in the strain-specific subclade B.Br.LVS is included within the subclade immediately basal to its actual subclade, B.Br.013/014. Colors within the mapped pie charts correspond to the subclade color designations. Gray regions indicate the known distribution of *F. tularensis*, by country (19). For explanation of stars and circles, see the legend to Fig. 3.

sis subclades contained greater MLVA diversity than *F. tularensis* subsp. *holarctica* subclades (Fig. 4 and 5, columns G). Considerable variation in MLVA genetic diversity was observed among the *F. tularensis* subsp. *holarctica* subclades, with subclades B.Br.OSU18 and B.Br.013/014 displaying the greatest number of *F. tularensis* subsp. *holarctica* MLVA genotypes (Fig. 5, column G).

Geographic distribution of clades and subclades. Figures 4 and 5 depict the geographic distribution of each of the subclades. North America, Sweden, and France are well represented in this study, with scattered representatives from other locations. Despite these sampling biases, major genetic and geographic trends are still apparent.

The subclades identified within the A.I and A.II subpopulations (two and four, respectively) showed no additional geographic associations, just the previously described eastern North American/Californian and western North American mountain trends for the A.I and A.II subpopulations, respectively (Fig. 4) (11, 19, 29). The two subclades A.I.Br.SCHU S4 and A.I.Br.001/002 are found throughout the eastern United States and in California. Subclade A.I.Br.001/002 is also found in Alaska and British Columbia, Canada, on the basis of the limited samples available from these locations (Fig. 4; also see Table S1 in the supplemental material). The presence of both of these subclades in California and the eastern United States is consistent with previous hypotheses that *F. tularensis* subsp. *tularensis* may have been introduced to these areas from the south central United States, where both A.I subclades are also found (Fig. 4) (11). These hypotheses are based upon a center-of-diversity argument wherein the geographic region with the greatest observed genetic diversity (the south central United States) is assumed to be the location where the tested group (A.I) originated, diversified, and then dispersed from. Two Slovakian isolates, the only known *F. tularensis* subsp. *tularensis* isolates from outside North America (14), also fell in subclade A.I.Br.SCHU S4 (see Table S1 in the supplemental material). This is consistent with previous reports regarding the close resemblance of these isolates to strain SCHU S4 (6, 18).

Subclades within subpopulation A.II exhibited slightly stronger geographic associations than those within subpopulation A.I. This is consistent with previous findings of a positive correlation between genetic and geographic distance for this

subpopulation (11). The basal subclade to the strain-specific subclade A.II.Br.ATCC 6223, A.II.Br.004/005, appears restricted to Utah, the origin of strain ATCC 6223 (Fig. 4; also see Table S1 in the supplemental material). Most of the subclade A.II.Br.003/004 isolates are from California, although single isolates from Nevada, Texas, and Ontario, Canada, also belong in this subclade (Fig. 4; also see Table S1 in the supplemental material). Finally, A.II.Br.001/002 and A.II.Br.006/007 (which includes subclade A.II.Br.WY96-3418) lack any strong geographic associations, being found throughout the geographic range of subpopulation A.II (Fig. 4).

We discovered 10 subclades within *F. tularensis* subsp. *holarctica*, most of which show geographic correlations (Fig. 5). The most basal *F. tularensis* subsp. *holarctica* subclade, B.Br.001/002, contains Japanese *F. tularensis* subsp. *holarctica* isolates. The next subclade, B.Br.002/003, contains two isolates from California and represents a previously unknown group within *F. tularensis* that is genetically distinct and diverged between the Japanese *F. tularensis* subsp. *holarctica* isolates and the terminal *F. tularensis* subsp. *holarctica* clade (Fig. 5; also see Table S1 in the supplemental material). The remaining *F. tularensis* subsp. *holarctica* subclades are closely related and likely reflect the radiation event (hereafter referred to as the B radiation) wherein *F. tularensis* subsp. *holarctica* dramatically spread throughout the northern hemisphere (Fig. 5). Subclade B.Br.FTNF002-00 corresponds to a previously described clone of *F. tularensis* subsp. *holarctica* found in France and the Iberian Peninsula (7). Interestingly, the subclade immediately basal to subclade B.Br.FTNF002-00, B.Br.010/011, contains the same Swedish isolate (F0228) (Fig. 5; also see Table S1 in the supplemental material) that was found to share RD_{23} with French and Iberian Peninsula isolates (7), making RD_{23} specific for branch B.Br.010 (Fig. 3). Subclades B.Br.012/013 and B.Br.013/014 (which includes subclade B.Br.LVS), marked by branch B.Br.012 (Fig. 3), are found in Sweden and Eurasia, respectively (Fig. 5). Subclades B.Br.008/009 and B.Br.OR96-0246, marked by branch B.Br.008 (Fig. 3), and subclade B.Br.OSU18 are found in North America (Fig. 5). Subclade B.Br.OSU18 is not specific to North America, however, because it contains seven Scandinavian isolates as well (Fig. 5; also see Table S1 in the supplemental material). Likewise, subclade B.Br.007/008, basal to subclades B.Br.008/009 and B.Br.OR96-0246, is found in Scandinavia (Fig. 5), making the three subclades identified by B.Br.007 (Fig. 3) nonexclusive for North America.

Many of these *F. tularensis* subsp. *holarctica* subclades correlate with groups previously identified using MLVA (see Table S1 in the supplemental material) (18). Specifically, B.Br.OSU18 corresponds to MLVA clade B.II; B.Br.012/013 and B.Br.LVS correspond to MLVA clade B.III; B.Br.007/008, B.Br.008/009, and B.Br.OR96-0246 (marked by branch B.Br.007) (Fig. 3) correspond to MLVA clade B.IV; B.Br.001/002 corresponds to MLVA clade B.V; and B.Br.010/011 and B.Br.FTNF002-00 (marked by branch B.Br.010) (Fig. 3) correspond to the Spain, France, and Sweden MLVA clade. Members of subclade B.Br.013/014 were found in two clades, B.I and B.III, in the previous MLVA (see Table S1 in the supplemental material) (18).

TABLE 2. *F. tularensis* canSNP genotypes[a]

Subclade	Nucleotide for indicated canSNP														
	M.Br.001	A/M.Br.001	A.Br.001	A.I.Br.001	A.I.Br.002	A.II.Br.001	A.II.Br.002	A.II.Br.003	A.II.Br.004	A.II.Br.006	B.Br.001	B.Br.002	B.Br.003	B.Br.004	B.Br.005	B.Br.006	B.Br.007	B.Br.008	B.Br.009	B.Br.010	B.Br.011	B.Br.012	B.Br.013
A.I.Br.001/002	C	A	T	A	G	C	C	T	C	C	G	C	G	A	C	G	T	G	A	A	G	A	A
A.I.Br.SCHU S4	C	A	T	A	A	C	C	T	C	C	G	C	G	A	C	G	T	G	A	A	G	A	A
A.II.Br.001/002	C	A	T	G	G	T	C	T	C	C	G	C	G	A	C	G	T	G	A	A	G	A	A
A.II.Br.003/004	C	A	T	G	G	T	T	C	C	C	G	C	G	A	C	G	T	G	A	A	G	A	A
A.II.Br.004/005	C	A	T	G	G	T	T	C	G	C	G	C	G	A	C	G	T	G	A	A	G	A	A
A.II.Br.006/007	C	A	T	G	G	T	T	T	C	T	G	C	G	A	C	G	T	G	A	A	G	A	A
B.Br.001/002	C	G	C	G	G	C	C	T	C	C	A	C	G	A	C	G	T	G	A	A	G	A	A
B.Br.002/003	C	G	C	G	G	C	C	T	C	C	A	G	G	A	C	G	T	G	A	A	G	A	A
B.Br.OSU18	C	G	C	G	G	C	C	T	C	C	A	A	G	A	C	G	T	G	A	A	G	A	A
B.Br.007/008	C	G	C	G	G	C	C	T	C	C	A	A	A	T	C	G	T	G	A	A	G	A	A
B.Br.008/009	C	G	C	G	G	C	C	T	C	C	A	A	A	A	T	A	C	T	A	A	G	A	A
B.Br.OR96-0246	C	G	C	G	G	C	C	T	C	C	A	A	A	A	T	A	C	T	G	A	G	A	A
B.Br.010/011	C	G	C	G	G	C	C	T	C	C	A	A	A	A	T	A	T	G	A	G	A	A	A
B.Br.FTNF002-00	C	G	C	G	G	C	C	T	C	C	A	A	A	A	T	A	T	G	A	G	A	A	A
B.Br.012/013	C	G	C	G	G	C	C	T	C	C	A	A	A	A	T	G	T	G	A	A	A	A	A
B.Br.013/014	C	G	C	G	G	C	C	T	C	C	A	A	A	A	T	G	T	G	A	A	G	T	G
M.Br.FSC 147	T	A	C	G	G	C	C	T	C	C	A	G	G	A	C	G	T	G	A	A	G	A	A
N	C	G	C	G	G	C	C	T	C	C	A	C	G	A	C	G	T	G	A	A	G	A	A

[a] This table lists the five synapomorphic terminal subclades and the 13 intervening subclades (Fig. 3, 4, and 5) together with their canSNP profiles. Each terminal subclade is defined by a single canSNP. Each intervening subclade is defined by two canSNPs, one on each side of the node.

DISCUSSION

F. tularensis is a highly successful pathogen with a distribution throughout the northern hemisphere (19). This wide geographic success is predominantly due to the spread of *F. tularensis* subsp. *holarctica*. However, very little genetic diversity has been identified within *F. tularensis* subsp. *holarctica* by using a variety of molecular methods (8, 11–13, 18, 19, 22, 29, 30), indicating that this subspecies only recently emerged through a genetic bottleneck and spread to its current distribution. The geographic origin of *F. tularensis* subsp. *holarctica* is debatable. Our data suggest that *F. tularensis* subsp. *holarctica* may have originated in Asia, as indicated by the basal positioning of FSC 022, the Japanese *F. tularensis* subsp. *holarctica* ("*F. tularensis* subsp. *japonica*") strain genome included in our whole-genome SNP analysis (Fig. 2). However, this clade is diverged from the main *F. tularensis* subsp. *holarctica* clade (Fig. 2) associated with the B radiation; while Asia may be the geographic origin of *F. tularensis* subsp. *holarctica*, it may not be the geographic origin of the main B radiation. An alternative hypothesis involves *F. tularensis* subsp. *holarctica* originating in North America. The basal positioning of the Californian subclade B.Br.002/003 in our MIP SNP analysis (Fig. 3 and 5) suggests that the main *F. tularensis* subsp. *holarctica* lineage either originated in or was introduced to North America prior to the divergence of this clade. If the former is the case, an early introduction to Asia from North America must have occurred to give rise to the Japanese *F. tularensis* subsp. *holarctica* isolates; if the latter is the case, then an early introduction to North America would have to have occurred. Regardless, given these data, parsimony would then suggest that the main *F. tularensis* subsp. *holarctica* lineage continued to diverge within North America and eventually gave rise to the highly successful *F. tularensis* subsp. *holarctica* clade that was spread around the northern hemisphere during the B radiation. The success of this B-radiation clade may be due to selectively advantageous differences between this clade and the more-basal B.Br.002/003 and Japanese *F. tularensis* subsp. *holarctica* clades.

In proposing a North American origin for the B radiation, one caveat must be mentioned. Specifically, it must be noted that we lack a comprehensive set of isolates from Asia. Given the possibility of an Asian origin for *F. tularensis* subsp. *holarctica*, as evidenced by the basal positioning of the Japanese *F. tularensis* subsp. *holarctica* isolates and our limited isolate coverage for Asia, it is possible that we are missing considerable diversity from Asia. Analysis of additional Asian isolates could reveal additional intermediate lineages that could preclude a North American origin for the B radiation. However, our current data set points to a North American origin, and we will discuss potential transmission patterns for the B-radiation clade on the basis of this hypothesis.

A North American origin for the B radiation is supported by the North American source of the subclade (B.Br.002/003) immediately basal to the divergence of the radiation clade (Fig. 5). This suggests that the main *F. tularensis* subsp. *holarctica* lineage was present in North America prior to the emergence of the B-radiation clade. The "Californian" B.Br.002/003 subclade appears to be a rare subclade that diverged from the main *F. tularensis* subsp. *holarctica* lineage after the divergence of the Japanese *F. tularensis* subsp. *holarctica* lineage but before the divergence of the ancestor to the B-radiation clade (Fig. 3 and 5). The presence of this subclade in California (Fig. 5) suggests that the main *F. tularensis* subsp. *holarctica* lineage was present in North America prior to the emergence of the B-radiation clade.

A North American origin for the B radiation is further supported by the likely North American source of the first subclade to diverge within the main B radiation. B.Br.OSU18, the first subclade to diverge within the main B radiation, contains both North American and Scandinavian isolates (Fig. 5; also see Table S1 in the supplemental material), but the basal B.Br.002/003 subclade argues for a North American rather than a Scandinavian emergence. In addition, the B.Br.OSU18 subclade is relatively rare in Scandinavia compared to other subclades (Fig. 5), and the relative lack of MLVA diversity in Scandinavian B.Br.OSU18 isolates compared to that in North American B.Br.OSU18 isolates suggests a recent introduction into the Old World. There were four MLVA genotypes among the seven Scandinavian B.Br.OSU18 isolates that differed only at the extremely diverse VNTR locus Ft-M03 (data not shown), a locus that has previously been shown to be diverse among even related outbreak isolates (18). This is in contrast to the North American B.Br.OSU18 isolates, which exhibited 22 MLVA genotypes among 45 unrelated isolates and differed at loci other than Ft-M03 (data not shown), indicating a higher level of diversity consistent with an older population. These frequency and diversity results are not likely due to sampling bias, given the relatively large sample sets analyzed from both North America and Sweden. Additional SNP discovery using one of these Scandinavian isolates and one or more North American B.Br.OSU18 isolates would confirm a North American origin for this subclade if subsequent SNP screening placed the North American B.Br.OSU18 isolates basally to the Scandinavian B.Br.OSU18 isolates. This, in turn, would support our hypothesis of a North American origin for the B-radiation clade.

Following a North American origin, the B-radiation clade would have to have spread and diversified throughout North America and been introduced to Eurasia, where it also must have spread, resulting in the current geographic distribution of *F. tularensis* subsp. *holarctica*. This clade may have spread throughout North America very rapidly, which would explain the wide distribution of subclade B.Br.OSU18 throughout North America (Fig. 5). However, the relatively large amount of MLVA diversity within this subclade (Fig. 5) suggests that additional geographic patterns likely exist. Additional whole-genome sequencing and SNP discovery efforts within subclade B.Br.OSU18 would likely reveal additional subclades. Mapping the geographic distributions of these novel subclades could provide additional insight into the spread of the B-radiation clade in North America and confirm or refute a rapid spread. A widespread basal subclade with geographically isolated distal subclades would indicate a rapid spread followed by local differentiation.

An introduction to Eurasia would have to have occurred and most likely is marked by branch B.Br.005 (Fig. 3), as nearly all of the Eurasian isolates are contained in the subsequent subclades (Fig. 5). Following its introduction to Eurasia, the B-radiation clade could have diverged and spread, giving rise

to subclades B.Br.007/008, B.Br.010/011, B.Br.FTNF002-00, B.Br.012/013, and B.Br.013/014 (Fig. 3 and 5). It is difficult to determine the exact distributions of these subclades due to the limited samples available from much of Europe and Asia. However, some patterns can be identified. Subclade B.Br.FTNF002-00 clearly represents a previously described clone (7) that was introduced to and spread throughout France and the Iberian Peninsula (Fig. 5). The lack of MLVA diversity among our expanded set of French B.Br.FTNF002-00 isolates (Fig. 5; also see Table S1 in the supplemental material) corroborates previous findings (7) and indicates that the spread of this clade throughout France and the Iberian Peninsula was likely a very recent event. The relative nearest to this clade, isolate F0228, is the only isolate within subclade B.Br.010/011 and may indicate a Scandinavian origin for the clade found in France and the Iberian Peninsula (Fig. 5; also see Table S1 in the supplemental material). Indeed, Scandinavia may be the origin point for many of the Eurasian subclades, as indicated by the canSNP diversity that we observed among the Scandinavian isolates. Subclades B.Br.007/008, B.Br.010/011, and B.Br.012/013 are exclusively found in Scandinavia, and subclade B.Br.013/014 is also present there (Fig. 5). This could easily be the result of sample bias, given the limited samples available from other European and Asian countries, but it may also suggest a Scandinavian origin for Eurasian *F. tularensis* subsp. *holarctica*. An expanded analysis of additional Central European and Asian isolates and additional SNP discovery could provide further insight into these possibilities. The presence of subclade B.Br.007/008, B.Br.010/011, or B.Br.012/013 in other locations could weaken the hypothesis of a Scandinavian origin for Eurasian *F. tularensis* subsp. *holarctica* and might suggest alternative scenarios for the spread of *F. tularensis* subsp. *holarctica* throughout Eurasia. Additional SNP discovery within subclade B.Br.013/014, the one subclade known to be present in Central Europe and Asia (Fig. 5), could also provide additional insight. Given the significant amount of MLVA diversity within subclade B.Br.013/014 (Fig. 5), the discovery of several additional canSNP groups is likely and could easily reveal additional geographic patterns and clues as to the spread of *F. tularensis* subsp. *holarctica* throughout Eurasia.

Additional, subsequent dispersals to both North America and Eurasia are required for a full explanation of the current distribution of *F. tularensis* subsp. *holarctica*. An additional introduction from North America to Scandinavia may have occurred, as evidenced by the seven Scandinavian isolates found in subclade B.Br.OSU18 (Fig. 5; also see Table S1 in the supplemental material). As discussed previously, these isolates likely reflect an introduction from North America to Scandinavia, rather than vice versa, given the small percentage of Scandinavian isolates belonging in this subclade (Fig. 5) and the greater MLVA diversity found among the North American B.Br.OSU18 isolates. Finally, there may have been a reintroduction event into North America from Eurasia, as evidenced by subclades B.Br.008/009 and B.Br.OR96-0246, which are the only North American groups in the Eurasian clade marked by branch B.Br.005 (Fig. 3 and 5). This reintroduction event is likely marked by branch B.Br.008, since it contains both subclade B.Br.008/009 and subclade B.Br.OR96-0246 (Fig. 3).

In summary, the clades, subclades, and lineages described in this study represent stable molecular groups that can be used to further classify *F. tularensis* global populations. A set of 23 canSNPs clearly defines four subspecies, two *F. tularensis* subsp. *tularensis* subpopulations, and 18 new synapomorphic subclades. As demonstrated here for *F. tularensis* subsp. *holarctica*, these groups provide an excellent means of tracking the origins and spread of this pathogen on a global scale. Discriminating among isolates within each of these groups for further epidemiological tracking requires the use of a higher-resolution molecular subtyping method, such as MLVA (18, 33). Combining these SNP and MLVA markers in progressive hierarchical resolving assays using nucleic acids, the approach suggested by Keim et al. (20), will provide highly accurate and highly discriminating phylogenetic analyses for *F. tularensis* where deeper phylogenetic relationships are defined by canSNP markers and strain discrimination is provided by MLVA. Subclades containing relatively high levels of MLVA diversity, such as B.Br.OSU18 and B.Br.013/014 (Fig. 5), indicate excellent isolate sets from which to select candidates for future whole-genome sequencing and SNP discovery. Sequencing isolates from these groups will provide the largest amount of new information useful for further elucidating the deeper phylogenetic structure of *F. tularensis*.

ACKNOWLEDGMENTS

This work was funded by the Department of Homeland Security Science and Technology Directorate, awards NBCH2070001 and HSHQDC-08-C-00158, and the Cowden Endowment in Microbiology at Northern Arizona University.

Note that the use of products/names does not constitute endorsement by the Department of Homeland Security of the United States.

We thank Sumathi Venkatapathy for technical assistance with the MIP assay.

REFERENCES

1. **Achtman, M., G. Morelli, P. Zhu, T. Wirth, I. Diehl, B. Kusecek, A. J. Vogler, D. M. Wagner, C. J. Allender, W. R. Easterday, V. Chenal-Francisque, P. Worsham, N. R. Thomson, J. Parkhill, L. E. Lindler, E. Carniel, and P. Keim.** 2004. Microevolution and history of the plague bacillus, *Yersinia pestis*. Proc. Natl. Acad. Sci. USA **101:**17837–17842.
2. **Auerbach, R. K.** 2006. sDACS: a novel *in silico* SNP discovery and classification method for bacterial pathogens. Northern Arizona University, Flagstaff, AZ.
3. **Beckstrom-Sternberg, S. M., R. K. Auerbach, S. Godbole, J. V. Pearson, J. S. Beckstrom-Sternberg, Z. Deng, C. Munk, K. Kubota, Y. Zhou, D. Bruce, J. Noronha, R. H. Scheuermann, A. Wang, X. Wei, J. Wang, J. Hao, D. M. Wagner, T. S. Brettin, N. Brown, P. Gilna, and P. S. Keim.** 2007. Complete genomic characterization of a pathogenic A.II strain of *Francisella tularensis* subspecies *tularensis*. PLoS ONE **2:**e947.
4. **Centers for Disease Control and Prevention, Office of Inspector General, Department of Health and Human Services (HHS).** 2005. Possession, use, and transfer of select agents and toxins. Final rule. Fed. Regist. **70:**13293–13325.
5. **Cha, R. S., H. Zarbl, P. Keohavong, and W. G. Thilly.** 1992. Mismatch amplification mutation assay (MAMA): application to the c-H-*ras* gene. PCR Methods Appl. **2:**14–20.
6. **Chaudhuri, R. R., C. P. Ren, L. Desmond, G. A. Vincent, N. J. Silman, J. K. Brehm, M. J. Elmore, M. J. Hudson, M. Forsman, K. E. Isherwood, D. Gurycová, N. P. Minton, R. W. Titball, M. J. Pallen, and R. Vipond.** 2007. Genome sequencing shows that European isolates of *Francisella tularensis* subspecies *tularensis* are almost identical to US laboratory strain Schu S4. PLoS ONE **2:**e352.
7. **Dempsey, M. P., M. Dobson, C. Zhang, M. Zhang, C. Lion, C. B. Gutiérrez-Martín, P. C. Iwen, P. D. Fey, M. E. Olson, D. Niemeyer, S. Francesconi, R. Crawford, M. Stanley, J. Rhodes, D. M. Wagner, A. J. Vogler, D. Birdsell, P. Keim, A. Johansson, S. H. Hinrichs, and A. K. Benson.** 2007. Genomic deletion marking an emerging subclone of *Francisella tularensis* subsp. *holarctica* in France and the Iberian Peninsula. Appl. Environ. Microbiol. **73:**7465–7470.
8. **Dempsey, M. P., J. Nietfeldt, J. Ravel, S. Hinrichs, R. Crawford, and A. K. Benson.** 2006. Paired-end sequence mapping detects extensive genomic re-

arrangement and translocation during divergence of *Francisella tularensis* subsp. *tularensis* and *Francisella tularensis* subsp. *holarctica* populations. J. Bacteriol. **188:**5904–5914.

9. **Easterday, W. R., M. N. Van Ert, T. S. Simonson, D. M. Wagner, L. J. Kenefic, C. J. Allender, and P. Keim.** 2005. Use of single nucleotide polymorphisms in the plcR gene for specific identification of *Bacillus anthracis*. J. Clin. Microbiol. **43:**1995–1997.
10. **Easterday, W. R., M. N. Van Ert, S. Zanecki, and P. Keim.** 2005. Specific detection of *Bacillus anthracis* using a TaqMan mismatch amplification mutation assay. BioTechniques **38:**731–735.
11. **Farlow, J., D. M. Wagner, M. Dukerich, M. Stanley, M. Chu, K. Kubota, J. Petersen, and P. Keim.** 2005. *Francisella tularensis* in the United States. Emerg. Infect. Dis. **11:**1835–1841.
12. **Fey, P. D., M. M. Dempsey, M. E. Olson, M. S. Chrustowski, J. L. Engle, J. J. Jay, M. E. Dobson, K. S. Kalasinsky, A. A. Shea, P. C. Iwen, R. C. Wickert, S. C. Francesconi, R. M. Crawford, and S. H. Hinrichs.** 2007. Molecular analysis of *Francisella tularensis* subspecies *tularensis* and *holarctica*. Am. J. Clin. Pathol. **128:**926–935.
13. **García Del Blanco, N., M. E. Dobson, A. I. Vela, V. A. De La Puente, C. B. Gutiérrez, T. L. Hadfield, P. Kuhnert, J. Frey, L. Domínguez, and E. F. Rodríguez Ferri.** 2002. Genotyping of *Francisella tularensis* strains by pulsed-field gel electrophoresis, amplified fragment length polymorphism fingerprinting, and 16S rRNA gene sequencing. J. Clin. Microbiol. **40:**2964–2972.
14. **Gurycová, D.** 1998. First isolation of *Francisella tularensis* subsp. *tularensis* in Europe. Eur. J. Epidemiol. **14:**797–802.
15. **Hardenbol, P., J. Banér, M. Jain, M. Nilsson, E. A. Namsaraev, G. A. Karlin-Neumann, H. Fakhrai-Rad, M. Ronaghi, T. D. Willis, U. Landegren, and R. W. Davis.** 2003. Multiplexed genotyping with sequence-tagged molecular inversion probes. Nat. Biotechnol. **21:**673–678.
16. **Hardenbol, P., F. Yu, J. Belmont, J. Mackenzie, C. Bruckner, T. Brundage, A. Boudreau, S. Chow, J. Eberle, A. Erbilgin, M. Falkowski, R. Fitzgerald, S. Ghose, O. Iartchouk, M. Jain, G. Karlin-Neumann, X. Lu, X. Miao, B. Moore, M. Moorhead, E. Namsaraev, S. Pasternak, E. Prakash, K. Tran, Z. Wang, H. B. Jones, R. W. Davis, T. D. Willis, and R. A. Gibbs.** 2005. Highly multiplexed molecular inversion probe genotyping: over 10,000 targeted SNPs genotyped in a single tube assay. Genome Res. **15:**269–275.
17. **Hopla, C. E., and A. K. Hopla.** 1994. Tularemia, p. 113–126. *In* G. W. Beran (ed.), Handbook of zoonoses, 2nd ed. CRC Press, Boca Raton, FL.
18. **Johansson, A., J. Farlow, P. Larsson, M. Dukerich, E. Chambers, M. Byström, J. Fox, M. Chu, M. Forsman, A. Sjöstedt, and P. Keim.** 2004. Worldwide genetic relationships among *Francisella tularensis* isolates determined by multiple-locus variable-number tandem repeat analysis. J. Bacteriol. **186:**5808–5818.
19. **Keim, P., A. Johansson, and D. M. Wagner.** 2007. Molecular epidemiology, evolution, and ecology of *Francisella*. Ann. N. Y. Acad. Sci. **1105:**30–66.
20. **Keim, P., M. N. Van Ert, T. Pearson, A. J. Vogler, L. Y. Huynh, and D. M. Wagner.** 2004. Anthrax molecular epidemiology and forensics: using the appropriate marker for different evolutionary scales. Infect. Genet. Evol. **4:**205–213.
21. **Larsson, P., P. C. Oyston, P. Chain, M. C. Chu, M. Duffield, H. H. Fuxelius, E. Garcia, G. Hälltorp, D. Johansson, K. E. Isherwood, P. D. Karp, E. Larsson, Y. Liu, S. Michell, J. Prior, R. Prior, S. Malfatti, A. Sjöstedt, K. Svensson, N. Thompson, L. Vergez, J. K. Wagg, B. W. Wren, L. E. Lindler, S. G. Andersson, M. Forsman, and R. W. Titball.** 2005. The complete genome sequence of *Francisella tularensis*, the causative agent of tularemia. Nat. Genet. **37:**153–159.
22. **Larsson, P., K. Svensson, L. Karlsson, D. Guala, M. Granberg, M. Forsman, and A. Johansson.** 2007. Canonical insertion-deletion markers for rapid DNA typing of *Francisella tularensis*. Emerg. Infect. Dis. **13:**1725–1732.
23. **Li, B., I. Kadura, D. J. Fu, and D. E. Watson.** 2004. Genotyping with TaqMAMA. Genomics **83:**311–320.
24. **Olsufjev, N. G., and I. S. Meshcheryakova.** 1983. Subspecific taxonomy of *Francisella-tularensis*. Int. J. Syst. Bacteriol. **33:**872–874.
25. **Papp, A. C., J. K. Pinsonneault, G. Cooke, and W. Sadée.** 2003. Single nucleotide polymorphism genotyping using allele-specific PCR and fluorescence melting curves. BioTechniques **34:**1068–1072.
26. **Pearson, T., J. D. Busch, J. Ravel, T. D. Read, S. D. Rhoton, J. M. U'Ren, T. S. Simonson, S. M. Kachur, R. R. Leadem, M. L. Cardon, M. N. Van Ert, L. Y. Huynh, C. M. Fraser, and P. Keim.** 2004. Phylogenetic discovery bias in *Bacillus anthracis* using single-nucleotide polymorphisms from whole-genome sequencing. Proc. Natl. Acad. Sci. USA **101:**13536–13541.
27. **Petrosino, J. F., Q. Xiang, S. E. Karpathy, H. Jiang, S. Yerrapragada, Y. Liu, J. Gioia, L. Hemphill, A. Gonzalez, T. M. Raghavan, A. Uzman, G. E. Fox, S. Highlander, M. Reichard, R. J. Morton, K. D. Clinkenbeard, and G. M. Weinstock.** 2006. Chromosome rearrangement and diversification of *Francisella tularensis* revealed by the type B (OSU18) genome sequence. J. Bacteriol. **188:**6977–6985.
28. **Rohmer, L., C. Fong, S. Abmayr, M. Wasnick, T. J. Larson Freeman, M. Radey, T. Guina, K. Svensson, H. S. Hayden, M. Jacobs, L. A. Gallagher, C. Manoil, R. K. Ernst, B. Drees, D. Buckley, E. Haugen, D. Bovee, Y. Zhou, J. Chang, R. Levy, R. Lim, W. Gillett, D. Guenthener, A. Kang, S. A. Shaffer, G. Taylor, J. Chen, B. Gallis, D. A. D'Argenio, M. Forsman, M. V. Olson, D. R. Goodlett, R. Kaul, S. I. Miller, and M. J. Brittnacher.** 2007. Comparison of *Francisella tularensis* genomes reveals evolutionary events associated with the emergence of human pathogenic strains. Genome Biol. **8:**R102.
29. **Staples, J. E., K. A. Kubota, L. G. Chalcraft, P. S. Mead, and J. M. Petersen.** 2006. Epidemiologic and molecular analysis of human tularemia, United States, 1964–2004. Emerg. Infect. Dis. **12:**1113–1118.
30. **Svensson, K., P. Larsson, D. Johansson, M. Byström, M. Forsman, and A. Johansson.** 2005. Evolution of subspecies of *Francisella tularensis*. J. Bacteriol. **187:**3903–3908.
31. **Van Ert, M. N., W. R. Easterday, L. Y. Huynh, R. T. Okinaka, M. E. Hugh-Jones, J. Ravel, S. R. Zanecki, T. Pearson, T. S. Simonson, J. M. U'Ren, S. M. Kachur, R. R. Leadem-Dougherty, S. D. Rhoton, G. Zinser, J. Farlow, P. R. Coker, K. L. Smith, B. Wang, L. J. Kenefic, C. M. Fraser-Liggett, D. M. Wagner, and P. Keim.** 2007. Global genetic population structure of *Bacillus anthracis*. PLoS ONE **2:**e461.
32. **Van Ert, M. N., W. R. Easterday, T. S. Simonson, J. M. U'Ren, T. Pearson, L. J. Kenefic, J. D. Busch, L. Y. Huynh, M. Dukerich, C. B. Trim, J. Beaudry, A. Welty-Bernard, T. Read, C. M. Fraser, J. Ravel, and P. Keim.** 2007. Strain-specific single-nucleotide polymorphism assays for the *Bacillus anthracis* Ames strain. J. Clin. Microbiol. **45:**47–53.
33. **Vogler, A. J., D. Birdsell, D. M. Wagner, and P. Keim.** 2009. An optimized, multiplexed multi-locus variable-number tandem repeat analysis system for genotyping *Francisella tularensis*. Lett. Appl. Microbiol. **48:**140–144.
34. **Vogler, A. J., E. M. Driebe, J. Lee, R. K. Auerbach, C. J. Allender, M. Stanley, K. Kubota, G. L. Andersen, L. Radnedge, P. L. Worsham, P. Keim, and D. M. Wagner.** 2008. Assays for the rapid and specific identification of North American *Yersinia pestis* and the common laboratory strain CO92. BioTechniques **44:**201, 203–204, 207.
35. **Whipp, M. J., J. M. Davis, G. Lum, J. de Boer, Y. Zhou, S. W. Bearden, J. M. Petersen, M. C. Chu, and G. Hogg.** 2003. Characterization of a *novicida*-like subspecies of *Francisella tularensis* isolated in Australia. J. Med. Microbiol. **52:**839–842.
36. **Worobey, M.** 2005. Anthrax and the art of war (against ascertainment bias). Heredity **94:**459–460.